

RESEARCH ARTICLE SUMMARY

MICROBIOTA

Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans

Sarah Moraïs, Sarah Winkler, Alvah Zorea, Liron Levin, Falk S. P. Nagies, Nils Kapust, Eva Lamed, Avital Artan-Furman, David N. Bolam, Madhav P. Yadav, Edward A. Bayer, William F. Martin, Itzhak Mizrahi*

INTRODUCTION: Mammals, including humans, rely on their gut's microbial community to break down plant cell wall components, notably cellulose and associated polysaccharides. However, there is limited evidence for cellulose fermentation in the human gut despite the benefits of cellulose-containing dietary fiber for gut-microbiome health and overall human well-being.

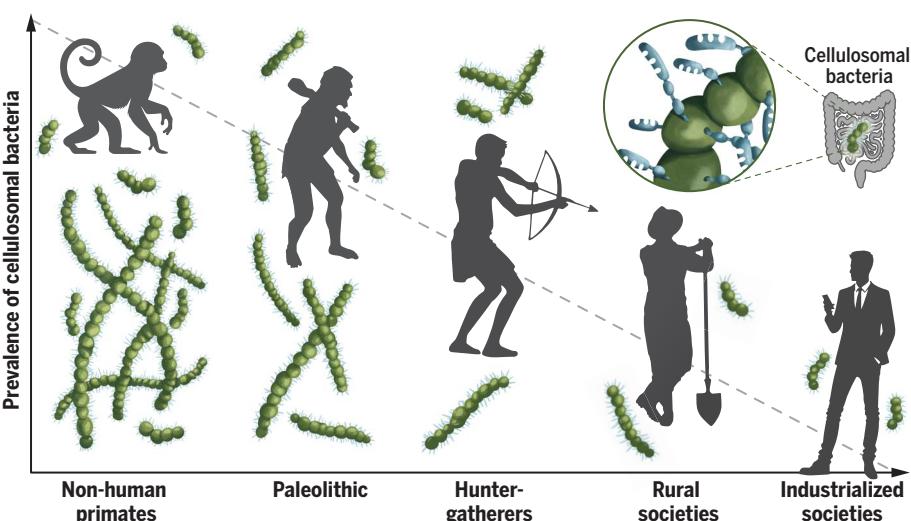
RATIONALE: By investigating the presence of heretofore undescribed bacterial species within the human-gut microbiota that degrade complex cellulosic polysaccharides, we can reveal their potential sources and understand their intricate adaptations to diverse host lifestyles and diets. Insight into the prevalence and abundance of these bacteria across diverse mammalian species and a wide range of human populations will provide critical knowledge of their evolutionary origins, ancestral associations, and trajectories that enabled their incorporation into the human gut.

RESULTS : Previously unknown ruminococcal species were discovered in the human-gut mi-

crobiota and provisionally named *Candidatus Ruminococcus primaciens*, *Ruminococcus hominiciens*, and *Ruminococcus ruminiciens*, all of which assemble functional multienzymatic cellulosome systems that degrade crystalline cellulose. These species are prevalent among the great apes and other nonhuman primates, ancient human societies, hunter-gatherer communities, and rural populations. Although widespread geographically they are conspicuously rare within industrialized societies. Notably, they exhibit distinct host preferences wherein *R. hominiciens* is associated primarily with humans and great apes and *R. primaciens* predominantly inhabits the gut of nonhuman primates and ancient human populations. Moreover, these species display host-specific diversification, forming distinct clades within the phylogenetic tree and aligning with their respective hosts. Our evolutionary analysis strongly suggests that *R. hominiciens* likely originated in the ruminant gut and later transferred to humans, possibly during domestication. High gene expression levels were observed for these species, reflecting their considerable

activity in their respective gut systems. Furthermore, their gene expression profile aligns with their hosts' dietary preferences, highlighting their adaptability. Our analyses show that these novel species adapt to their host ecosystems by acquiring genes from co-resident gut microbes. The human-associated strains possess functional adaptability highlighted by the acquisition of genes that can degrade specific plant fibers of monocots such as maize, rice, and wheat—major components of the human diet. Likewise, the nonhuman primate-associated strain exhibits the potential for degrading chitin, a polymer abundant in the insect exoskeleton, part of the diet of nonhuman primates. Our data provide insight into the ongoing colonization of these species within the human gut, particularly those originating from ruminants and nonhuman primates. Specific strains appear to represent intermediates between primate- and rumen-gut ecosystems, as evidenced by their gene content during establishment in the human intestine.

CONCLUSION: Our accumulated data indicate that ruminococcal lineages were more widespread in the past, evidenced by the high prevalence and abundance of these strains in ancient human populations and among hunter-gatherer communities and rural societies, combined with their global distribution and low prevalence in industrialized societies. Differences in their prevalence among human populations may reflect dietary variation between industrialized and nonindustrialized societies. Dietary fiber intake appears to be a key factor as high-fiber diets are reported among Hadza hunter-gatherers whereas lower fiber intake is observed in rural populations and the least consumption of fiber occurs in industrialized societies. These findings collectively imply a decline of these species in the human gut, likely influenced by the shift toward westernized lifestyles, potentially impacting energy balance and other health-related aspects. The presence of transitional strains that recently colonized the human gut indicates that ruminants and nonhuman primates could be a source and reservoir for cellulosome-producing ruminococcal strains, which continue to colonize and adapt to the human gut. There may be potential for intentional reintroduction or enrichment of these species in the human gut through targeted dietary approaches and specialized probiotics. ■



Cellulose degrading gut bacteria of hominids across evolutionary time. Previously unknown human gut cellulolytic ruminococcal species are highly prevalent in nonhuman primates, the great apes, ancient human populations, hunter-gatherer communities, and in rural populations but are rare in urbanized human populations.



The list of author affiliations is available in the full article online.

*Corresponding author. Email: imizrahi@bgu.ac.il

Cite this article as S. Moraïs et al., *Science* **383**, eadj9223 (2024). DOI: 10.1126/science.adj9223

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.adj9223>

RESEARCH ARTICLE

MICROBIOTA

Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans

Sarah Moraïs^{1,2,3}, Sarah Winkler^{1,2,3}, Alvh Zorea^{1,2,3}, Liron Levin⁴, Falk S. P. Nagies⁵, Nils Kapust⁵, Eva Lamed⁶, Avital Artan-Furman⁶, David N. Bolam⁷, Madhav P. Yadav⁸, Edward A. Bayer^{1,2,6}, William F. Martin⁵, Itzhak Mizrahi^{1,2,3*}

Humans, like all mammals, depend on the gut microbiome for digestion of cellulose, the main component of plant fiber. However, evidence for cellulose fermentation in the human gut is scarce. We have identified ruminococcal species in the gut microbiota of human populations that assemble functional multienzymatic cellulosome structures capable of degrading plant cell wall polysaccharides. One of these species, which is strongly associated with humans, likely originated in the ruminant gut and was subsequently transferred to the human gut, potentially during domestication where it underwent diversification and diet-related adaptation through the acquisition of genes from other gut microbes. Collectively, these species are abundant and widespread among ancient humans, hunter-gatherers, and rural populations but are rare in populations from industrialized societies thus indicating potential disappearance in response to the westernized lifestyle.

Dietary fiber is beneficial to gut microbiome stability and richness and has important implications for human health (1). Fermentation of dietary fiber in the human gut regulates digestive transit, prevents obesity and diabetes, and reduces cardiovascular diseases and cancer (1). Microbial activity transforms these indigestible glycans into short-chain fatty acids which supply energy to the host and have multiple effects not only on the gut but also systemically (2). Cellulose is a major part of the plant cell wall (3) and consequently a common component of diets that include plant-based components. The benefits of cellulose on host health have been shown in animals and include prevention of colon cancer (4) and reduction in blood sugar levels (5). The prevalence of cellulose in processed food is very low but there is a growing preference to decrease the amount of processed food ingredients in favor of a plant-based diet with increased fiber levels.

It was long believed that crystalline cellulose was not digested in the human gut, in contrast to ruminants and other herbivores (6, 7). Evidence for the degradation of microcrystalline cellulose—the purified crystalline cellulose portion from cellulose fibers—by human gut bacteria was first reported in 2003 (8) and the microcrystalline cellulose degrader *Ruminococcus chamanellensis* was isolated a decade later (9). Subsequently, the presence of cellulosomes—multi-enzymatic complexes that degrade plant-fiber polysaccharides—were detected in this bacterium. Biochemical characterization of its interactive cellulosomal proteins and enzymes confirmed its full functionality (10–12). Despite this discovery, cellulose degradation and fermentation in the human gut is rare or absent in most humans (13, 14). Nevertheless, the presence of cellulosomes across gut ecosystems indicates that they play a distinct role in promoting energy release from dietary fiber.

Despite considerable progress, fundamental questions remain concerning the prevalence of cellulosome-producing bacterial species in the mammalian gut, their adaptability to host lifestyle and diet, and whether other undiscovered cellulosome-producing bacterial species reside in the human gut. In this study we aimed to address these questions. We used the human strain *R. chamanellensis* and the related rumen species *Ruminococcus flavefaciens* as reference cellulosome-producing bacterial species (5, 6, 9) to identify related species by searching for key cellulosome genes in metagenome-assembled genomes. We examined the functionality of the cellulosomes in the species we discovered, how these functions are rooted within these bacterial lineages, their connection to their respective host lifestyles and diets, and the dynamics

of their evolutionary trajectory from our primate relatives to diverse human cultures.

This group of human gut bacteria produce functional cellulosomes, are phylogenetically related to the rumen-based *R. flavefaciens*, and are prevalent in several nonhuman primate (NHP) lineages. We found that these bacteria have diversified within their various host ecosystems and have adapted to their lifestyles by acquiring genes from their surrounding microbial communities. These cellulosome-carrying species are prevalent at low incidence in westernized human populations but at higher levels in ancient human, hunter-gatherer, and non-westernized societies. Our data also indicate that strains of these ruminococcal species are continuing to colonize the human gut from NHPs and ruminants and are dynamically adapting to the human gut ecosystem.

Results

Detection of fiber-degrading species in the human gut microbiome

By identifying known cellulosomal components in genomes of *Ruminococcus* spp., we aimed to determine the breadth of the diversity of human gut cellulosome-producing species. Cellulosome complexes are heterogeneous modular assemblies of structural proteins (scaffoldins) and enzyme arrays that target different recalcitrant plant fiber components (Fig. 1A). The cellulosome complex is composed of multiple scaffoldins that contain a multiplicity of cohesin modules each of which interacts with a complementary dockerin module located on each of the cellulosomal enzyme components (Fig. 1A).

To retrieve and analyze cellulosome-producing ruminococcal genomes we used the *scaC* gene that encodes a definitive cellulosomal scaffoldin protein and that so far is known only in the *Ruminococcus* genus (15, 16) (Fig. 1A). Using this approach we searched for *ScaC* sequences in 4941 rumen metagenome-assembled genomes (MAGs) from domesticated ruminant cattle and 92,143 human MAGs (17, 18), and identified 251 ruminococcal genomes that contain *ScaC*. After filtering genomes exhibiting at least 90% completion as determined by CheckM (19), we obtained 25 and 22 genomes of rumen and human origin, respectively (table S1). Maximum likelihood phylogenetic analysis of their *ScaC* sequences revealed a clustering pattern that almost completely distinguishes between human and rumen clades of ruminococcal genomes. This analysis was augmented by *ScaC* sequences of 12 sequenced genomes of *R. flavefaciens* isolates from the rumen environment and three sequenced genomes from isolates of their close relative from the human gut, *R. chamanellensis*.

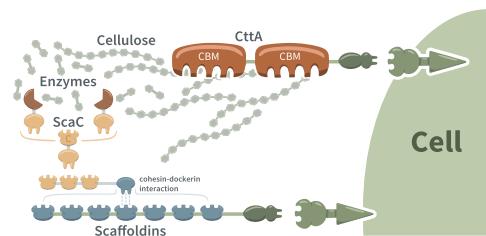
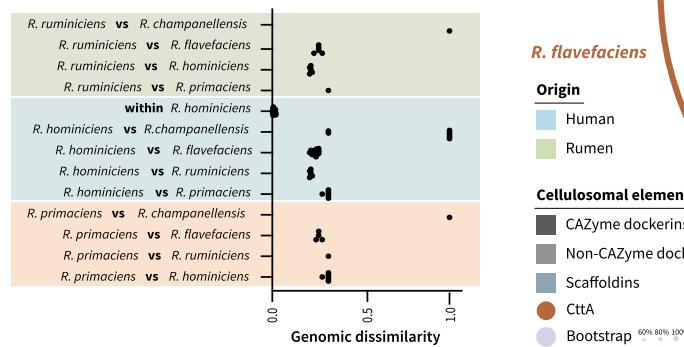
To deepen the phylogenetic analysis we further examined the fibrolytic potential of these 62 genomes by searching for the presence of

¹National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, 84105, Israel.

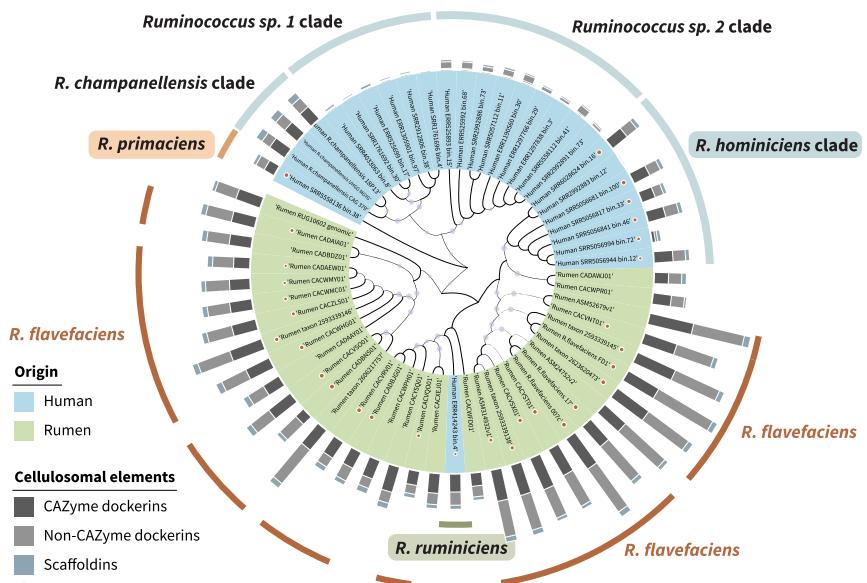
²Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel. ³The Goldman Sonnenfeld School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel.

⁴Bioinformatics Core Facility, Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Beer-Sheva, Israel. ⁵Department of Biology, Institute for Molecular Evolution, Heinrich-Heine-Universität Düsseldorf, D-40225, Düsseldorf, Germany. ⁶Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot 7610001 Israel. ⁷Institute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK. ⁸US Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center, 600 East Mermaid Lane, Wyndmoor, PA 19038, USA.

*Corresponding author. Email: imizrahi@bgu.ac.il

A The cellulosome**C Genomic comparison of strains****Fig. 1. Detection of a human-gut, fiber-degrading ruminococcal species.**

(A) Scheme of cellulosome architecture. The CttA protein by virtue of its CBMs mediates the binding of the bacterial cell to the cellulosic substrate which can be hydrolyzed by dockerin-bearing enzymatic units that are integrated into the cell-surface cellulosome through its cohesin-containing scaffoldin assemblies. **(B)** Unrooted phylogenetic tree computed with the maximum likelihood method of 62 selected genomes and MAGs using the sequence of the ScaC scaffoldin illustrated in Fig. 1A as a phylotyping marker (15, 16) (table S1). The color of the clade indicates the origin of the genomic bin (light blue, human; light green, rumen). Light purple circles on the

B Cellulosomal phylogeny and cellulosome content by host association

branches represent bootstrap values higher than 60%. The number and composition of cellulosomal elements is indicated as a bar for each genomic bin (number of dockerin-containing proteins with additional CAZyme elements, dark gray; number of dockerin-containing proteins with no additional CAZyme elements, medium gray; number of scaffolds containing at least one cohesin module, light gray). Brown circles next to the MAG name indicate genomes containing a *cttA* gene. **(C)** Genomic dissimilarity computed by Mash distance within the identified ruminococcal cellulosomal species and pairwise comparisons to each other as well as to the ruminal *R. flavefaciens* species and the human species *R. chamanellensis*.

cellulosomal elements and CAZymes (i.e., carbohydrate-active enzymes that act on glycosidic bonds) (20). We sought to identify the potential of enzyme components that integrate into cellulosome complexes, which would be detected by the presence of a dockerin module on the enzyme. We thus identified a total of 3687 dockerin-containing proteins among which 1853 also contained a CAZyme module (Fig. 1B), including glycoside hydrolases (GH), carbohydrate esterases, polysaccharide lyases, and carbohydrate-binding modules (CBMs) from various families. In addition, a total of 308 scaffolds were recovered. The phylogenetic clusters of the tree corresponded to the distribution of the functional cellulosomal components of the identified MAGs. The human-associated MAGs were separated into four distinct clades (bootstrap values higher than 90%) (Fig. 1B): two exhibited low numbers of cellulosomal elements (designated as *Ruminococcus* sp. 1 and *Ruminococcus* sp. 2 in the figure) whereas the remaining two exhibited high numbers of cellulosomal elements. The two latter clades were examined further and one was found to comprise sequences from

R. chamanellensis. Notably, the second contained ScaC sequences that were phylogenetically closer to those of the *R. flavefaciens* rumen isolate genomes (bootstrap value of 60%). The latter genomes also contained a *cttA* gene marker characteristic of the *R. flavefaciens* scaffoldin gene cluster. CttA is a cellulosomal protein that binds the bacterium to cellulose (Fig. 1A) (21). The gene for this cellulosome component represents a marker specific to *R. flavefaciens* that is absent from the human gut bacterium *R. chamanellensis*. The *cttA* gene can therefore be used specifically to distinguish between the two closely related cellulosome-producing species. Consequently, members of the clade that encode the *cttA* gene and occur in the human gut potentially represent additional human gut fiber-degrading cellulosomal species. We found an average of >99% similarity among this clade to each other but only 78% similarity to the genomes of isolates and MAGs affiliated with the rumen *R. flavefaciens* (Fig. 1C) (22). In addition, we retrieved the 16S-rRNA gene sequence of four of the six MAGs which were found to show an average of 95.8 and 92.7% identity to the

rumen *R. flavefaciens* and human *R. chamanellensis* species, respectively, and 100% identity to each other (table S2). This finding supported their potential association as a distinct ruminococcal species, which we registered as '*Candidatus "Ruminococcus hominicensis"* sp. nov.' in the SeqCode registry (23).

Two MAGs of human origin that also encoded the *cttA* gene marker and numerous cellulosomal elements were not located within the *R. hominicensis* clade. Our data for genome similarity and marker genes (specified below) showed that these MAGs may also represent distinct cellulosome-producing bacterial species occupying similar niches to *R. hominicensis* (Fig. 1B). One MAG was positioned within the rumen-associated MAG clade, and the second appeared as a single isolated branch of the phylogenetic tree. The 16S-rRNA sequence of the former MAG was not available but it exhibited low average genome similarities to the *R. hominicensis* (80%) and *R. flavefaciens* genomes (75.6%) (Fig. 1C, green background, and table S2). The latter MAG also exhibited low genome similarity to the *R. hominicensis* and *R. flavefaciens* strains, 71 and 77.3%, respectively (Fig. 1C, orange

background), and its 16S-rRNA sequence exhibited relatively low identity to the latter strains as well (90.6 and 91.3%, respectively). These data suggest that the strains are distinct species and thus were provisionally named with the SeqCode registry. The human-associated MAG that was positioned within the rumen clade was named '*Candidatus "Ruminococcus ruminicola"*' sp. nov.' and the other human-associated MAG that appeared as a single branch on the phylogenetic tree was named '*Candidatus "Ruminococcus primaciensis"*' sp. nov.' In addition, Protologger analysis (24) of the *R. ruminicola*, *R. primaciensis* and *R. hominicola* genomes indicated that these are species with potential for cellulose and starch utilization as well as acetate, propionate, and L-glutamate production, similar to that of *R. flavefaciens* (strain FD-1).

Fiber-degrading bacterial species prevalence in nonindustrialized humans

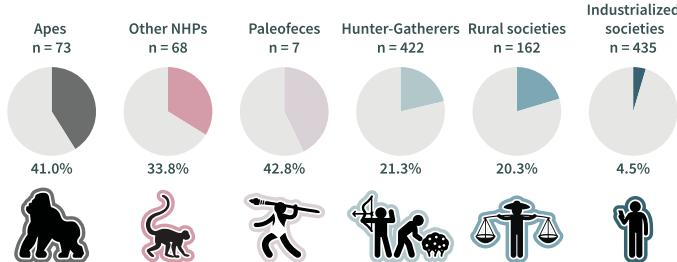
The prevalence and abundance of the fiber-degrading species and known ruminococcal

species, *R. flavefaciens* and *R. hominicola*, were investigated across 1989 gut samples of humans and animal species worldwide (Fig. 2A, fig. S1, and table S3). The samples originated from 75 animal species, including wild and domesticated animals (NHP and ruminants), as well as various human cohorts. This analysis revealed that the human-associated genotypes (*R. primaciensis*, *R. hominicola*, and *R. ruminicola*) are broadly distributed (Fig. 2B) and are specific to humans and several NHP species (i.e., macaques, baboons, gorillas, and chimpanzees), but absent from the ruminant samples tested (see figs. S1, S2, and S3). In addition, the rumen MAGs were specific to ruminants but absent from the human and NHP cohorts tested (figs. S1, S2, and S3).

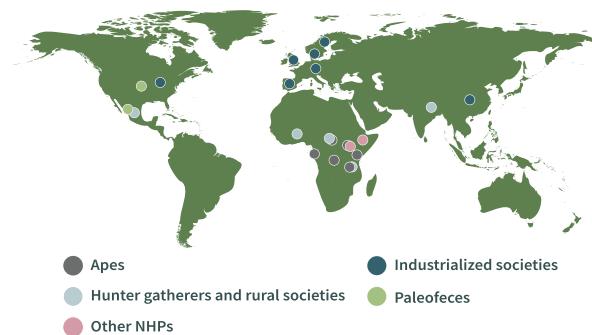
The prevalence and abundance of *R. primaciensis*, *R. hominicola*, and *R. ruminicola* displayed notable variations among diverse human cohorts. In industrialized countries, including Denmark, China, Sweden, and the USA, the collective prevalence of these strains

reached a maximum of 4.6% (Fig. 2A) with some notable differences in *R. hominicola* prevalence between these countries (fig. S4). All three strains exhibited higher collective prevalence in the different cohorts of the non-industrialized populations we tested: 43% prevalence in human paleofeces samples dating from 1000 to 2000 years ago (25), 21% in hunter-gatherers, and 20% in geographically diverse rural societies (with no significant differences among geographies, fig. S5). Samples from apes and other NHPs had 41% and 33% prevalence, respectively (Fig. 2A and fig. S2, A and B). Furthermore, the abundance of these strains in each positive individual was significantly lower in industrialized populations when compared with all nonindustrialized human populations, as well as in apes and other NHP samples (fig. S2C). The rumen strain was more abundant in ruminants than human strains for both human and NHP samples (fig. S6). The variations in prevalence of these species in human populations could potentially be linked

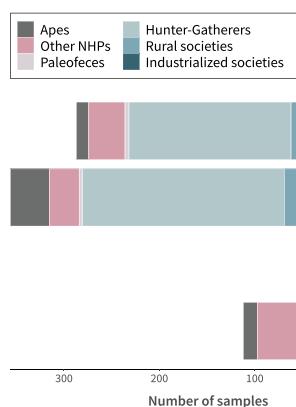
A Collective prevalence of novel fiber-degrading species



B Geography of novel fiber-degrading species' identification



C i. Host-strain association



ii. Distribution of strains across hosts

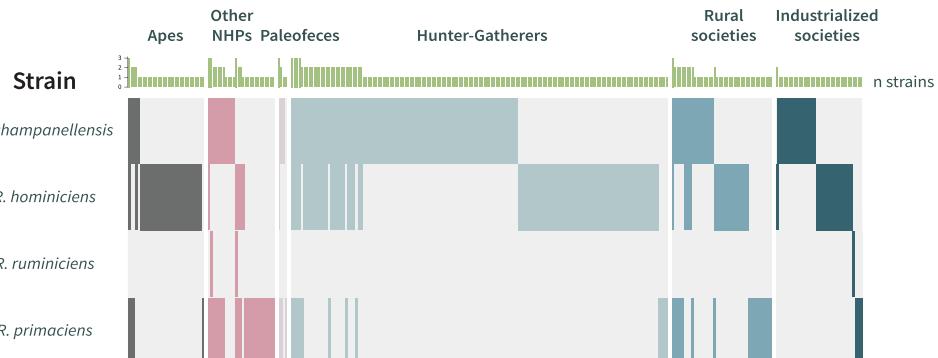


Fig. 2. *Ruminococcus* spp. are abundant in ancient human, hunter-gatherer, and rural populations. (A) Observed collective prevalence of the MAGs for fiber-degrading strains in various human, ape, and NHP cohorts. Pie charts represent the observed prevalences. (B) Worldwide locations of positive human and NHP samples. The locations of the samples in which the human MAGs were detected are denoted on the map as circles: dark blue, industrialized societies; light blue, rural societies and hunter-gatherers; green, paleofeces; pink, wild NHP. (C) Distribution of fibrolytic strains in human and NHP populations. (i) Stacked bar chart of the distribution of each human cellulosomal strain (*R. chamaenellensis*, *R. hominicola*, *R. ruminicola*, and *R. primaciensis*) across the sample cohorts. (ii) Heatmap of the distribution of the human cellulosomal strains among the human- and NHP-positive samples. The bar plot above the heatmap represents the number of strains detected in each sample.

and pink, wild NHP. (C) Distribution of fibrolytic strains in human and NHP populations. (i) Stacked bar chart of the distribution of each human cellulosomal strain (*R. chamaenellensis*, *R. hominicola*, *R. ruminicola*, and *R. primaciensis*) across the sample cohorts. (ii) Heatmap of the distribution of the human cellulosomal strains among the human- and NHP-positive samples. The bar plot above the heatmap represents the number of strains detected in each sample.

to dietary disparities between individuals in industrialized and nonindustrialized societies (26, 27), as well as human activities that affect microbial diversity such as the use of antibiotics (28). Dietary fiber intake may be a major contributing factor given its close association with the prevalence and abundance of these species. Notably, adult Hadza hunter-gatherers typically consume 80 to 150 g per day (30) of dietary fiber whereas rural populations have substantially lower estimates at 13 to 14 g per day (31, 32), and industrialized populations even less at 8.4 g per day. Moreover, the prevalence of *R. hominiciens* strains in wild versus captive apes was significantly lower in prevalence in captive animals further strengthening the connection between lifestyle and diet on the prevalence of these strains (fig. S7). In other NHP samples *R. primaciens* was more prevalent in omnivorous than in folivorous monkeys, suggesting that the fiber content in these diets is sufficient and that other factors may also play a role (fig. S8). Furthermore, the high prevalence and abundance of these strains in human samples dating back 1000 to 2000 years (25) and among hunter-gatherer populations, coupled with the global distribution of the human *Ruminococcus* spp. strains (Fig. 2B), suggests that although these lineages currently exist in limited proportions of human populations they were previously more widespread and abundant, consistent with a recent study that shows loss of taxa while humans speciated from great ape relatives and while switching from a non-industrialized to industrialized lifestyle (29).

We found similar levels of prevalence for the fiber-degrading species and the previously identified *R. chamaenellensis* cellulolytic strains in human gut samples (fig. S1), which led us to investigate the potential exclusion or cooperation processes that might drive the distribution of these species and strains. Analysis of the strain distribution of *Ruminococcus* spp. revealed that when fiber intake is high, as in nonindustrial countries, strain diversity increases whereas in most of the samples originating from humans of industrial countries, only one fibrolytic strain was detected, indicating potential competitive exclusion among these species when fiber intake is low (Fig. 2Cii). An alternative scenario to exclusion would be the stochastic effects of loss due to antimicrobial selection in industrialized countries. By contrast, human samples from either hunter-gatherer societies or nonindustrialized countries as well as apes and other NHP samples exhibited various combinations of two or more species of *Ruminococcus* spp., which suggests reduced competition possibly attributable to greater access to fiber-rich diets and/or increased niche availability. Niche availability such as carbohydrate diversity may enable niche partitioning among strains through variations in glycolytic hydrolysis-coding genes present in their genomes,

ultimately leading to a higher diversity of fibrolytic strains in these samples (Fig. 2Ci). The examination of different strains' prevalence and abundance within individual hosts allowed us to also investigate host-strain associations. Our findings provided evidence of distinct host preferences among the various strain lineages. Specifically, *R. primaciens* exhibited a significant association with other NHPs and ancient humans (indval test *P*-value = 0.01) whereas *R. hominiciens* is significantly associated with humans and apes (indval test *P*-value = 0.005; Fig. 2Cii). Furthermore, *R. ruminiciens*—characterized by its higher similarity to the rumen strains (see below)—was found to be rare in all samples (Fig. 2Ci).

Ongoing colonization by ruminococci in the human gut

We studied the potential evolutionary scenarios for core proteins found in all genomes of the ruminococcal strains. Because we have also identified these strains in NHPs, we augmented our MAG set with eight additional MAGs originating from NHP-gut samples (30). The latter genomes were assembled with at least 90% genome completion as analyzed by CheckM and are 98% similar to the *R. primaciens* strain. We predicted and clustered the overall open reading frames (ORFs) from the different strains' genomes (14 rumen-, 8 human-, and 8 NHP-associated MAGs) and clustered them into 5958 orthologous groups using the Proteinortho program (31). The different host-associated strains shared a core genome composed of 315 orthologous protein groups from which we generated maximum likelihood trees that were colored according to the samples in which the MAGs were assembled (fig. S9). In all of the trees the proteins were clustered according to the respective host (Fig. 3, A and B)—that is, human NHP and ruminant—suggesting within-host clonal diversification and potential speciation with the exception of two strains, one being a *R. primaciens* MAG and the second corresponding to a *R. ruminiciens* MAG, both assembled from human samples and thereby suggesting recent transfer from NHP and ruminants to the human gut (Fig. 3).

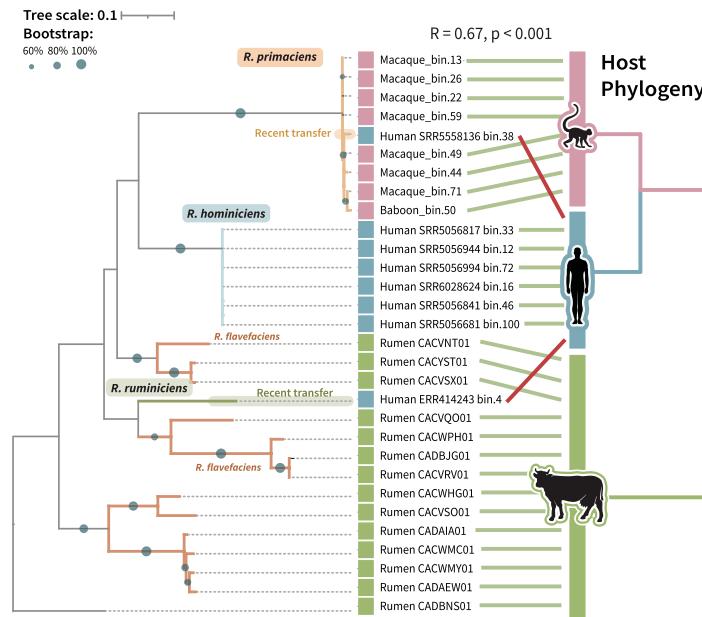
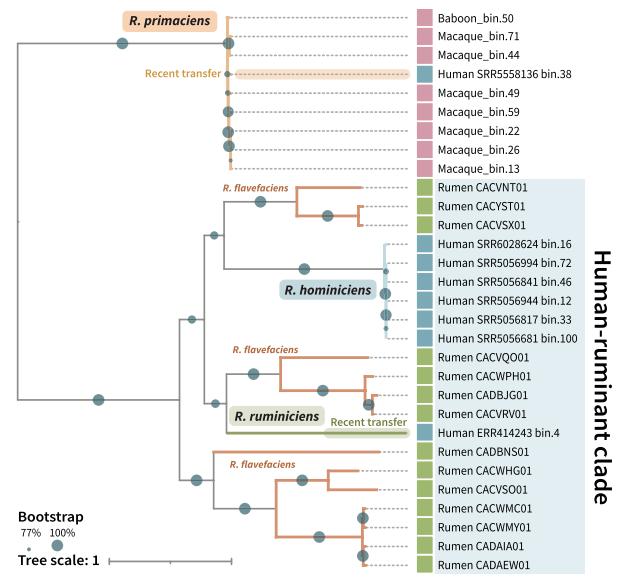
In some cases, a cospeciation scenario emerged, with host phylogeny significantly correlated with matching associated strains such as human hosts with human strains and primate strains with NHP hosts supported by both Mantel correlation and AU tests (Fig. 2A). However, in most of the trees, as well as in both an multilocus sequence analysis (MLSA) tree and a concatenated tree comprising the majority of core genes (fig. S10, A and B), the human-associated strain clade was closer to the ruminant than NHP clade (Fig. 3B).

In our ancestral analysis we traced the origin of *R. hominiciens* strains back to their roots in ruminant strains supported by a significant

92% bootstrap split in the concatenated tree that places the human associated clade within the ruminant clade. This pattern remained consistent across the majority of individual trees within both evolutionary scenarios (see fig. S10). Phylogenetic analysis of the concatenated tree demonstrated that *R. primaciens* strains associated with NHPs exhibited a significantly shorter phylogenetic distance to the ancestor of all human strains when compared with that of the *R. hominiciens* strains (one-sided Wilcoxon rank-sum test *P*-value: 0.000174) (Fig. 3B). Collectively these findings strongly suggest that *R. primaciens* is the closest relative to the ancestors of all human strains and that the ancestors of the *R. hominiciens* strains originated from ruminant strains. We can thus speculate that the transfer to humans occurred during the domestication process, with these strains subsequently adapting and diversifying within the human gut environment.

Functional cellulosome fiber-degradation and cellulose-adhesion activities

Our phylogenetic analyses identified *R. primaciens* strains as the closest to the ancestor of human cellulolytic strains and indicated recent transfer of this species into the human gut (Fig. 3). This discovery provided the opportunity to investigate whether the ancestral human gut *R. primaciens* strain can efficiently degrade crystalline cellulose and produce active assembled cellulosomes composed of components common to the other ruminococcal strains. To this end, we identified both scaffolds and enzymes that were shared among *R. primaciens*, *R. hominiciens*, *R. ruminiciens*, and *R. flavefaciens* (table S4). We examined their potential for cellulosome assembly using the matching fusion-protein approach (32) in which the binding abilities of the recombinant proteins—seven cohesin and six dockerin modules—were measured (table S5). Out of the 36 potential interactions tested, 10 positive interactions were thus identified which enabled us to predict the cellulosomal assembly of these modules (Fig. 4A and table S6). The proposed structure of the *R. primaciens* cellulosome (Fig. 4B) resembles the known *R. flavefaciens* cellulosomal organization in strains isolated from ruminants (33). In both *R. primaciens* and *R. flavefaciens* strains, the scaffoldin proteins show a similar interaction pattern whereby the dockerins of the ScaA and ScaC scaffolds interact with the cohesins of ScaB through divergent cohesin-dockerin interactions (see Fig. 4A). The cellulosome is attached to the microbial cell wall through selective cohesin-dockerin interaction between ScaB and ScaE. Furthermore, the dockerin-containing enzymes interact with their cohesin counterparts of ScaA, ScaB, and ScaC with divergent specificities. Finally, similar to ScaB, CttA is integrated into the bacterial cell wall by means of

A Co-speciation hypothesis**B Human-ruminant clade in a concatenated tree of 197 core proteins****Fig. 3. Colonization by ruminococci is ongoing and dynamic in the human gut.**

(A) Core protein phylogenetic tree illustrating the cospeciation hypothesis (left panel). Blue circles on the branches represent bootstrap values higher than 60%. The comparison with the phylogenetic tree of the mammalian host species is given on the right with red lines indicating proteins that do not recapitulate host phylogeny.

(B) Phylogenetic tree of 197 concatenated core proteins. Blue circles on the branches

represent bootstrap values higher than 77%. Blue highlighting on the right indicates a close phylogenetic distance between the human and ruminant clades. In (A) and (B) MAGs are color-coded according to host origin: green, blue, or pink indicate rumen, human, or NHP, respectively; transitional strains are denoted as "recent transfers" and the tree scales represent the number of amino acid substitutions per site. MAGs corresponding to *Ruminococcus flavefaciens* are indicated.

a similar type of cohesin-dockerin interaction with ScaE. We measured the ability of cellulosomal components from the two species as well as from *R. chamaeleonensis* to interact with each other. We found cross-species interactions of cellulosomal components of *R. primaciens* with representative cohesin-dockerin combinations from *R. chamaeleonensis* 18P13 and *R. flavefaciens* FD-1, indicating evolutionary conservation of the interaction residues and a certain degree of promiscuity among these components (tables S7, S8, and S9).

We selected one of the GH5 cellulase enzymes for biochemical characterization of its cellulolytic activity as this type of GH5 gene was common to 25 of the 30 MAGs used in our analyses. The GH5 enzyme exhibited cellulolytic activity on microcrystalline cellulose as a substrate (Fig. 4C and fig. S11) and its enzymatic activity was in a range similar to that of the *R. flavefaciens* FD-1 ortholog (68% sequence identity). We also purified the CtaA protein from *R. hominiciens* and found that it exhibited robust binding to microcrystalline cellulose (Fig. 4D) indicating that the bacterial cells would bind to cellulose owing to the interaction with the cell-wall-anchored scaffoldin ScaE (see below and Fig. 4B) (34, 35). Altogether these results demonstrate that the cellulosomes of the ruminococcal strains are assembled and active on the crystalline cellulose substrate.

Specific host gut adaptation

The phylogenetic clustering of *R. hominiciens*, *R. primaciens*, and *R. ruminiciens* strains according to their hosts (Figs. 1B and 3, A and B), along with the significant association of *R. hominiciens* to humans and apes and of *R. primaciens* to other NHPs and ancient humans, raise the question of whether host association is reflected in the coding capacity of the different strains. The genomes from the different host ecosystems (14 rumen, 8 human, and 8 NHP MAGs) showed host specificity in their gene content and expression pattern, in accordance with their respective host's dietary preferences.

Principal component analysis (PCA) of the 5958 orthologous groups obtained earlier (Fig. 5A and fig. S9) showed host specificity in the genome content of the ruminococci we identified, yielding three distinct clusters corresponding to the different hosts (PERMANOVA $P < 0.001$), with the exception of two human assembled MAGs for *R. primaciens* and *R. ruminiciens*, which were located in the NHP clade and rumen clade, respectively (Fig. 5A). These results further support the notion that these strains represent a transitional adaptation stage. We further analyzed all strains for their core and flexible host-associated genomes to track the evolutionary trail that potentially brought about host adaptivity. Our analysis favored gene acquisition from external lineages

as the more probable scenario that allowed these lineages to adapt to different hosts.

The different host-associated strains shared a core genome composed of 315 orthologous groups common to the three species and a total of 233 host-specific orthologous groups that were found in all genomes of the given host-associated strains but not in the others (rumen, human, or NHP; fig. S9). We therefore asked to what degree the host-specific genes are rooted within the strain lineage as compared with the core genes. To this end we applied verticality analysis that measures the degree by which core and the host-specific genes are rooted within a strain phylogeny (36). While comparing verticality values for core proteins to those for host-specific orthologous groups, we found significantly higher values for the former (Fig. 5B). This finding indicated that host-specific genes were most probably gained by these strains from microbes that were coinhabiting the same specific host-associated gut environment whereas the core genes are endogenous to these strains and rooted within their lineage.

The identified ruminococcal species are suspected to occupy the fiber-degrading niche within gut ecosystems and their prevalence correlates with the dietary fiber content of their hosts (Fig. 2, A and C). Hence, genome adaptivity to the host environment should also be apparent in the gene composition of the fiber-degrading functions. We therefore analyzed

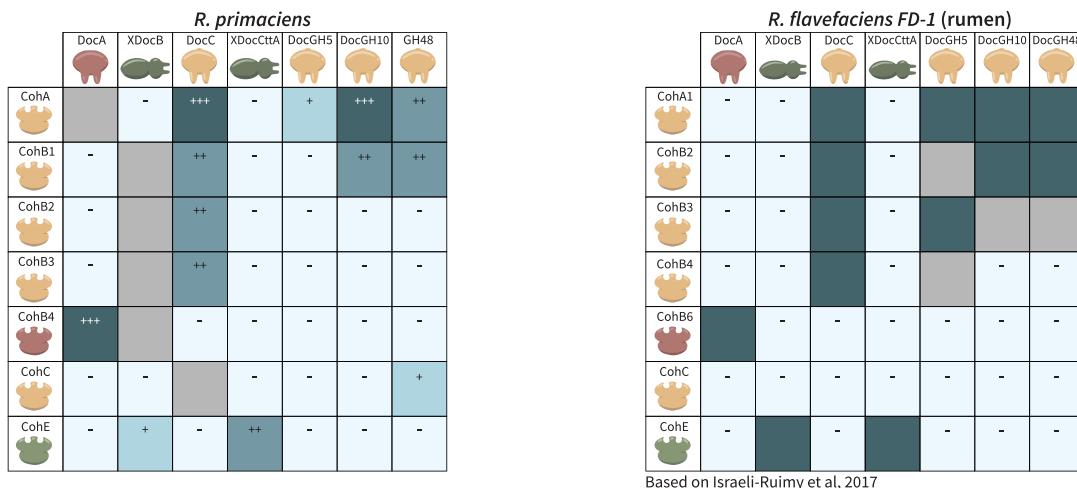
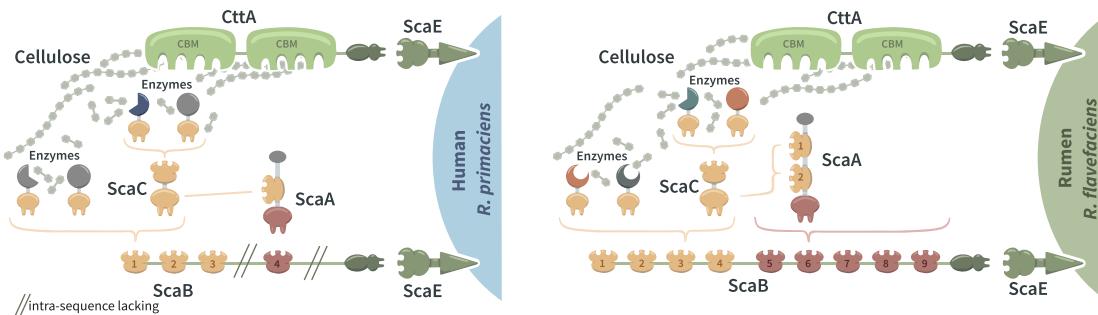
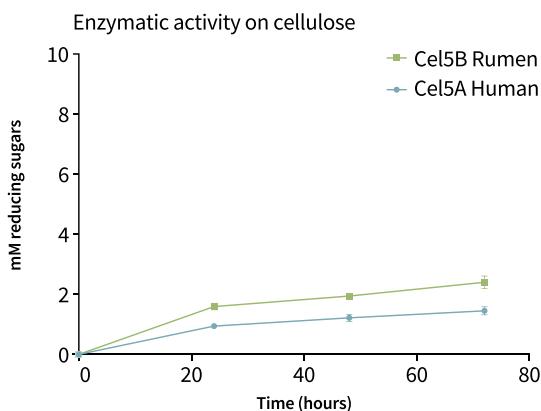
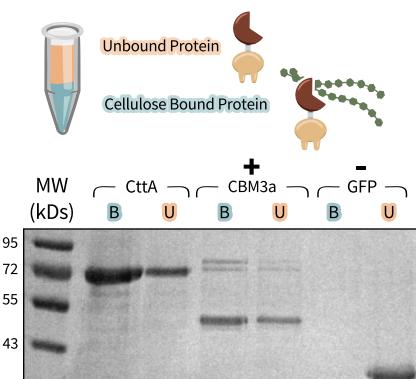
A Cellulosomal component interactions of human and rumen strains**B Schematic representation of cellulosome assembly in human and rumen strains****C Comparison of cellulase orthologs of human and rumen strains**

Fig. 4. Cellulosome assembly activity and cellulose adhesion. (A) Summary of interactions between selected cellulosomal recombinant cohesin and dockerin modules derived from an *R. primaciens* strain (Human_SRR5558136_bin.38) compared with those of orthologous modules from the *R. flavefaciens* FD-1 rumen strain (79). Cohesin and dockerin modules are color-coded (red, yellow, or green) according to their predicted specificities of interaction. On both panels, light blue highlights negative interactions; darker blue, positive interactions; gray, not tested. On the left panel (*R. primaciens*), intensities of the interactions are denoted with – for no affinity, (OD_{450} lower than 0.15), + for moderate affinity (OD_{450} between 0.15 and 0.5), ++ for high affinity (OD_{450} between 0.5 and 1.0), and +++ for very high affinity (OD_{450} between 1.0 and 2.2), respectively. On the right panel (*R. flavefaciens*), intensities were not available

D Human strain CttA-mediated bacterial cellulose-adhesion

for the Israeli-Ruimy 2017 study. (B) Overview of cellulosomal interactions in *R. primaciens* compared with those of *R. flavefaciens* as deduced from affinity-based ELISA experiments and proposed recognition residues of the dockerin components (table S6). (C) Comparative cellulolytic activity of ruminococcal GH5 orthologs of either human (*R. primaciens*) or rumen origin (*R. flavefaciens* FD-1). Enzyme samples were examined using microcrystalline cellulose (Avicel) as the substrate at 37°C. The data points represent the average of biological triplicates with standard deviation. (D) Cellulose binding assay. SDS-PAGE gels loaded with cellulose-bound (B) and -unbound (U) fractions of either *R. hominici* CttA, the CBM3a from the CipA scaffoldin of the *Clostridium thermocellum* cellulosome as a positive control or green fluorescent protein (GFP) as a negative control (nonbinding protein).

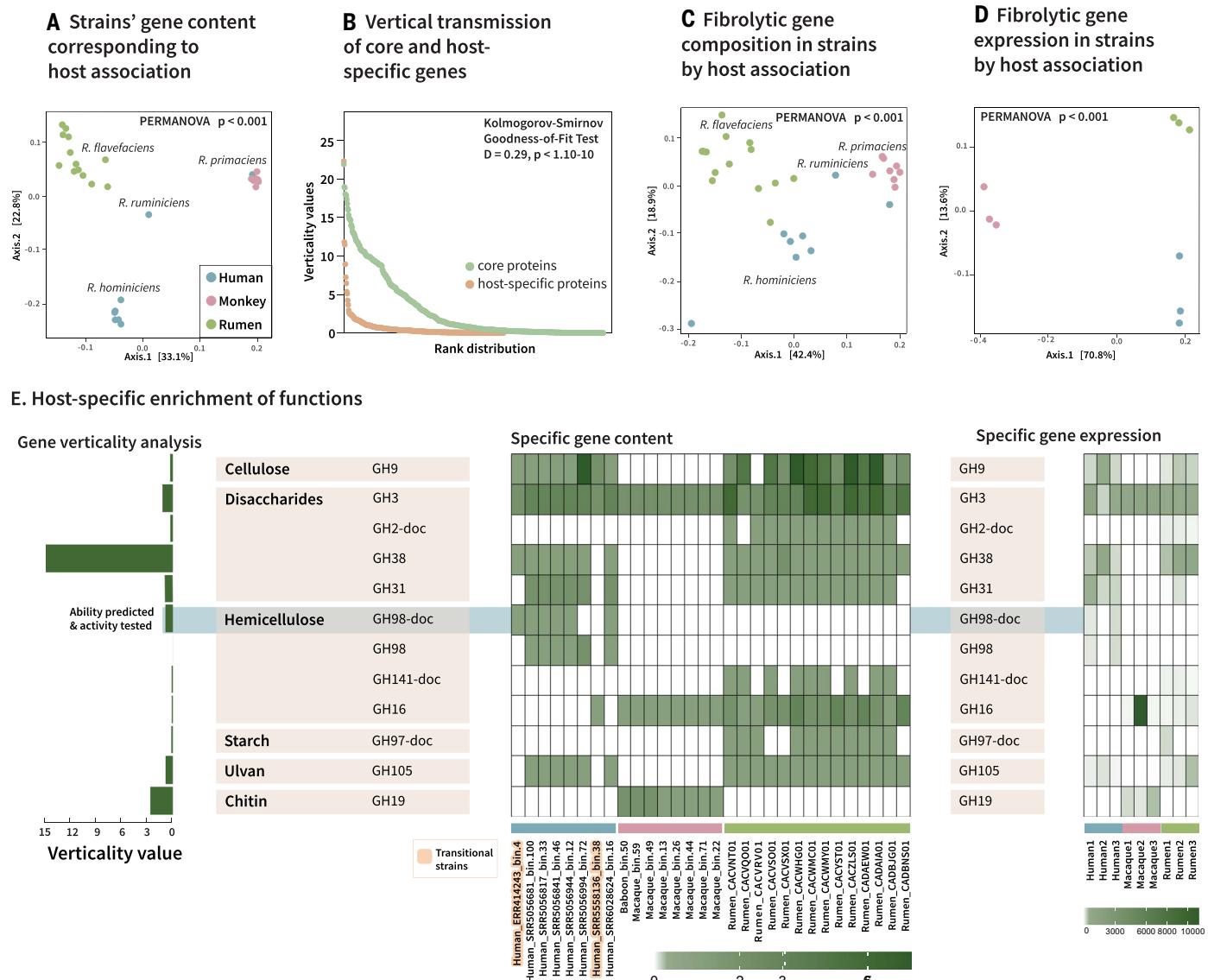


Fig. 5. Functional adaptation of MAGs with their host. In (A), (C), (D), and (E), MAGs and samples are color-coded according to host origin: green, blue, or pink indicating rumen, human, or NHP, respectively. (A) Principal component analysis (PCA) of the overall predicted ORFs of the MAGs, color-coded by their hosts (see below). Clustering analysis of MAG gene content according to their hosts was performed using the PERMANOVA test with 1000 randomizations of the data and the *P*-value is indicated. (B) Rank distribution of verticality values for core proteins across the three host types versus host-specific proteins indicates that specific genes are likely to be transferred through horizontal gene transfer within a given type of host. (C) PCA of the fibrolytic system [indicating glycoside hydrolase (GH) families] of the MAGs color-coded by their hosts. Clustering analysis of MAGs GH family content according to their hosts was

performed using PERMANOVA test with 1000 randomizations of the data and the p-value is indicated. (D) PCA of the expression of the fibrolytic system as examined by transcriptomic analysis of three fecal samples of the three hosts (macaque, human, and sheep rumen). (E) Center panel: heatmap of the statistically significant GH families that distinguish the strains associated with the three gut ecosystems as determined by the Kruskal-Wallis test $P < 0.05$ after FDR correction. The left bar graph represents the verticality values for each of these orthologous groups of genes. (Right) heatmap of the statistically significant GH expression (metatranscripts in FPKM) between the three types of hosts (see material and methods section). For the GH141-Doc and GH97-Doc genes, the metatranscripts were aligned to Rumen_CADBJG01 and Rumen_CACVQ001 MAG sequences (59).

the repertoire of fiber-degrading enzymes from these strains (table S7). We found that the glycoside hydrolase (GH) families coded by the different cellulosomal strains grouped into distinct clusters on a PCA plot according to their host, further corroborating host adaptation for fiber degradation (Fig. 5C).

Metatranscriptome data of three samples from each host gut ecosystem was analyzed by read alignment to the strain's genome and showed that these fiber-degrading genes are expressed within their host gut ecosystems pointing to high activity in the respective gut systems (fig. S12A). In all samples of the three

hosts, expression of 50 to 82% of their overall gene content was observed (fig. S12A). When examining only cellulosomal genes, even higher ratios were obtained with more than 90% of cellulosomal gene expression of *R. flavefaciens*, *R. hominiciens*, and *R. primaciens* in sheep, humans, and NHPs (fig. S12, B to D). These

include a variety of key fibrolytic functions and specific cardinal cellulases (GH5, GH9, and GH48) and hemicellulases (GH10, GH11, and GH26) that are mutual to these strains (figs. S12 and S13). In addition, the amount and function of cellulosomal gene expression between the triplicate samples from the same ecosystem were almost identical, indicating the presence of a specific realized niche of fiber degradation for these bacteria within each of the host gut environments (fig. S12, B to D). Although high similarity exists at the cellulosomal gene content and its expression level between the strains, the fine-tuned differences in gene presence and absence that are related to cellulosomal adaptation to the different ecosystems were also apparent at the expression profile (Fig. 5D).

By analyzing the fiber-degrading gene repertoire of the different species using the Kruskal Wallis test we highlighted specific GH families that statistically distinguish the strains associated with the three gut ecosystems (Fig. 5E and table S3). These findings showed that within the different host gut ecosystems there are specific host-related dietary components that trigger expression of these host-specific genes. For example, dockerin-containing GH families 2, 97, and 141 were only present in the *R. flavefaciens* rumen-associated strains and absent from the human- and NHP-gut *R. hominiciens* and *R. primaciens* genomes. These enzymes encode various hemicellulolytic activities such as mannosidase, glucoamylase, and xylanase activities, thus attesting to the richer spectrum of polysaccharides that exists in the rumen environment. Similarly, GH families coding for enzymes acting on cellulose (GH3 and GH9), mannans (GH31 and GH38), or arabinogalactan (GH105) were specific or present in higher numbers in both *R. hominiciens* and *R. flavefaciens* genomes and absent from *R. primaciens* genomes (table S7). In general, rumen-associated *R. flavefaciens* genotypes are richer in GH diversity and gene copy number than the *R. hominiciens* genomes, both of which were richer compared with that of *R. primaciens* (table S7). Collectively, these differences could be related to the notion that rumen strains participate in the degradation of a major substrate critical to host survival and that the rumen system provides higher retention times whereas the human-based strains reside in the colon and deal with the undigested remnants of what has already passed through most the digestive tract with shorter retention time.

Two GH families tightly connected to the host dietary constraints were found to be coded and expressed exclusively within host-associated strain: GH family 19, which includes putative chitinases and was exclusive to NHP-associated MAGs of *R. primaciens*, and GH family 98 which includes arabinoxylanases and was

exclusive to *R. hominiciens* genomes. Notably, in the MAGs for which we hypothesize transitional stages of adaptation—i.e., the human-associated MAGs of *R. primaciens* and *R. ruminiciens*—the GH98 gene is either lacking or present in only one copy respectively (Fig. 5E), which further suggests that these MAGs are in the process of adaptation to the human host. Likewise the GH19 gene is absent in the human MAG of *R. primaciens* which could suggest the loss of this function in human hosts.

These host-exclusive functions could be explained by host diet as the GH19 family found in *R. primaciens* genomes retrieved from NHP samples includes putative chitinases, which would presumably serve to degrade chitin of the insect exoskeleton ingested by the NHPs. GH98 enzymes found exclusively in the *R. hominiciens* genomes would potentially hydrolyze glucuronoarabinoxylan, a hemicellulose that constitutes 25% of the primary cell walls of monocots such as rice, wheat, and maize, which are major components of the modern human diet (37). To test this we further cloned and purified the putative GH98 enzyme of *R. hominiciens* and measured its ability to degrade corn glucuronoarabinoxylan as a model substrate (fig. S14 and Fig. 5E, left) thus confirming the potential role of GH98 in the adaptation of the human-associated *R. hominiciens* strain to the host diet.

Like other host-specific genes, these host-exclusive functions all have extremely low verticality values (0.004, 0.86, and 2.61 for GH98, GH98-Doc, and GH19, respectively), which suggests potential transmission to the human and NHP strains through horizontal gene transfer from the respective gut ecosystem (Fig. 5E, left graph). Indeed, the putative GH98 catalytic modules exhibited 44% sequence identity to the GH98 enzyme of *Bacteroides ovatus*, which was characterized as a glucuronoarabinoxylanase and potentially could be acquired from this lineage (38).

Discussion

We have identified three distinct, heretofore undescribed, cellulosome-producing, cellulytic human gut ruminococcal species: *Candidatus R. hominiciens*, *R. primaciens*, and *R. ruminiciens*. Our evolutionary analysis strongly suggests that *R. primaciens* is the closest strain to the common ancestor of all human strains and that *R. hominiciens* likely originated in the ruminant gut and was later transferred to humans, possibly during domestication. Nevertheless, cospeciation cannot be ruled out at this time. These species underwent diversification and host adaptation in their respective gut ecosystems. Notably, host adaptation of these strains primarily occurs through gene acquisition from other members of the microbiome, as demonstrated by verticality analysis.

These species appear to be declining in the industrialized human gut. Nevertheless, comprehensive understanding of their impact will be attained by future isolation of these strains and investigation of their physiology, fiber degradation potential, and effects on the host.

The presence of these microbes in the human gut can offer significant benefits within the context of subsistence diets by maximizing nutrition from locally available foods in resource-limited societies, potentially providing energy through metabolic products. Indeed, these gut microbes are scarce in industrialized populations but thrive in hunter-gatherer and rural communities where processed food consumption is minimal, and accompanied by a higher intake of natural, unprocessed plant fiber. Additionally, these microbes are highly prevalent and abundant in primates and in 1000 to 2000-year-old human gut samples, thus suggesting that they may have been an integral part of the ancestral human microbiome, consistent with a recent study that reported a higher prevalence of *R. champanellensis* in ancient and nonindustrialized human gut microbiomes (25).

Our research has revealed that these species continue to actively invade the human gut, as particularly evident in the case of strains of *R. primaciens* and *R. ruminiciens*. Although found in the human gut, their genomes appear to represent intermediates between primate- and rumen-gut ecosystems as they establish themselves in the human intestine, indicating that ruminants and NHPs may act as a source and reservoir for important cellulosome-producing ruminococcal strains, which continue to colonize and adapt to the human gut ecosystem. In this regard, a potential exists for their re-introduction or enrichment in the human gut through targeted diets and specialized probiotics.

Materials and Methods

Retrieval and analysis of ruminococcal genomes containing cellulosomal elements

The ScaC sequence from *R. flavefaciens* strain FD-1 (accession number CAK18894) was used as a query sequence to retrieve metagenome-assembled genomes (MAGs) of rumen and human origin (17, 18), using local blast (39). Hits below E-values of 10^{-4} , above 45% of sequence identities and of lengths higher than 250 amino acids were retained. Among these, only associated MAGs with above 90% completeness as determined by CheckM (19) were analyzed further. ScaC sequences were aligned using MegaX (40). Annotation of glycoside hydrolases in the selected genomes were performed with dbcan2 (41). The presence of the N-terminal sequence of the CttA protein (21) (427 amino-acids, accession number CAK18897.1), which corresponds to the cellulose-binding component of the cellulosome system,

was used as a specific marker for *R. flavefaciens* strains using tblastx.

Analysis of selected MAGs

Dockerin and cohesin-containing sequences were retrieved from the predicted proteome [using Prokka (42)] as detailed by Phitsuwan *et al.* (43). Annotation of dockerin-containing genes was performed using dbcan2. Mash analysis on the nucleotide level was performed on the genomes annotated using CttA as a marker (44).

Prevalence of selected MAGs in rumen and gut samples

At first, the 30 selected MAGs of rumen, human and NHP origin were aligned to their original sample reads (table S10). The number of reads were normalized between samples, and only alignments above 80% completion were retained. A heatmap of MAG abundances in the different samples was created, using the superheat package (<https://CRAN.R-project.org/package=superheat>). Then, to examine the prevalence of selected MAGs across gut samples from human and animals, we clustered the different MAGs that contained the CttA marker (Fig. 1B) based on 97% similarity, using the drep algorithm (45). This step resulted in 3 human and 8 rumen MAGs representing the three human gut species (*R. primaciens*, *R. hominiciens* and *R. ruminicola*) and various strains of *R. flavefaciens*. The MAGs were aligned to metagenomes from gut or rumen fecal samples (25, 27, 29, 30, 46–68). Samples with coverage of at least 20% for a given MAG at a threshold of 1 were considered positive. To normalize the variation in read depth between metagenomes, each metagenome was subsampled to 5, 10, 20, 40, and 60 million reads and each MAG prevalence was assessed as stated previously. A cutoff of 10 million reads was determined optimal for comparative analysis. Prevalence for *R. chamanellensis* was calculated similarly by aligning the 18P13 genome to the same fecal samples.

Evolutionary analysis of the selected MAGs

Proteinortho (31) was used to group orthologous proteins from human, rumen and NHPs genomes. For each of the 315 orthologous groups comprising the core genome shared between the different host-associated strains, a phylogenetic tree was created using the minimal ancestor deviation (MAD) rooting approach (69). Moreover, we searched for orthologs in the genome of *Clostridium thermobutyricum* DSM 4928 to serve as an outgroup. Outgroup orthologs were retrieved for 197 orthologous groups, and phylogenetic trees were created using the iqtree2 program package with 1000 bootstraps (70). We then performed an approximately unbiased (AU) analysis (71) on all core proteins for

which outgroup orthologs were available (197 core proteins out 315) to test a cospeciation scenario. We used, as a hypothesis scenario, one of the core protein trees that exhibited a high and significant correlation to the mammalian host's evolutionary tree, [using the dendext R package with cor.dendlist function (correlation of 0.67, *P*-value <0.001) (Fig. 3A)] (72). The AU test was performed as part of the iqtree2 program package (70) while using the '-au' parameter as well as the '-zb 10,000' parameter to indicate the number of RELL (73) replicates to perform several tree topology tests for all 197 core orthologous groups trees. We then performed an host/parasite cospeciation test [using the 'hommola_cospeciation' function from the 'skbio' python package (74)] similar to Sanders *et al.* (29) to identify core protein trees that exhibited similar host clustering as the mammalian host's evolutionary tree [created using the Timetree database (75)]. We also used the Mantel test (using the 'mantel.rtest' function from the 'ade4' R package), which yielded similar results to the hommola cospeciation test. We concatenated all the 197 core orthologous groups proteins and created a phylogenetic tree using iqtree2 program package with 1000 bootstraps (70). To examine whether *R. primaciens* is significantly closer to the most recent common ancestor of all strains identified in humans, we calculated the distance of each strain to the outgroup in the concatenated tree. We used the Wilcoxon rank sum exact test (two-sided) to test whether the distances of *R. primaciens* to the outgroup are smaller than the distances of all other human strains. All data and code are available in GitHub repository (76).

To perform MLSA (77), amino-acid sequences of the subunit of RNA polymerase (rpoB), subunit of DNA gyrase (gyrB), translation initiation factor IF-2 (infB), RNA modification GTPase ThdF or TrmE (thdF), chaperonin GroEL (groEL) and sigma 70 (sigma D) factor of RNA polymerase (rpoD) were retrieved from each of the 30 MAGs, aligned, concatenated using MegaX (40) and a maximum likelihood phylogenetic tree was generated.

Cloning of cellulosomal modules and enzymes from human strains

Thirteen sequences of dockerins and cohesins were selected from the *R. primaciens* strain and synthesized by IDT (Coralville, Iowa, USA) with additions of restriction sites at both ends. The synthesized DNA sequences of cohesins and dockerins were inserted into CBM-Coh and Xyn-Doc plasmid cassettes respectively (32), using appropriate restriction endonucleases (ThermoFisher Scientific). T4 ligase (New England Biolabs) was used for plasmid ligation and *Escherichia coli* strain DH5 alpha (Bio Lab, Israel) was used for transformation. Plasmids were verified by Sanger sequencing.

The sequence of a GH5 enzyme from *R. primaciens* strain was also synthesized by IDT and cloned into pET28a, using either restriction or restriction-free cloning. The N-terminal sequence of the GH5 was reconstructed using the consensus sequence of highly similar GH5 sequences, recovered by blastp (fig. S15). GH98 was cloned from metagenomic DNA extracted using the phenol-chloroform method (78) from a human sample, in which the CttA gene was detected using specific primers for CttA (table S11), cleaved using NcoI and XhoI and inserted into restricted pET28a by ligation. The list of all primers used in this study is available in table S11. The amino-acid sequences of the proteins used in the study are available in table S5.

Expression and purification of recombinant proteins and GH-containing dockerins

The proteins were expressed and purified as described earlier (17) with incubation at 37°C for 3 hours following induction with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The Xyn-Doc fusion proteins and GH-containing dockerins were purified using Ni-NTA beads (EMD, MERCK-Millipore) and CBM-Coh fusion proteins using amorphous cellulose (PASC).

Affinity-based ELISA analysis of cohesins using immobilized dockerins

The procedure of Barak *et al.*, was followed (32). Cohesins and dockerins from *R. chamanellensis* 18P13 and *R. flavefaciens* FD-1 for cross-species interactions were cloned and produced as described earlier (10, 79). All binding affinity assays were performed at least twice in biological triplicates.

Enzymatic activity assay

Cellulolytic activity was tested with 0.5 μM of either GH5 from *R. primaciens* or from *R. flavefaciens* FD-1 (table S5) on 1% Avicel microcrystalline cellulose (FMC, Delaware USA) at pH 5 (50 mM acetate buffer, final concentration) for 24, 48, and 72 hours at 37°C. Kinetics of amorphous cellulose degradation were followed by incubating the GH5 enzymes at concentrations ranging from 0 to 1 μM at pH 5 for 1 hour at 37°C with 7.5 g/l substrate. After incubation, the tubes were centrifuged for 2 min at 14,000 rpm at room temperature, and 100 μL of supernatant fluids were added to 150 μL dimethylsulfoxide (DMSO) solution (80), boiled for 10 min, and the absorbance at 540 nm was measured. Released sugar concentrations were determined using a glucose standard curve.

Glucuronoarabinoxylanase activity was tested by incubating 0.2% corn glucuronoarabinoxylan (38) in 20 mM citrate buffer (pH 6) with 20 μL of either purified GH98, double-distilled water (ddw) or the lysate of a *R. flavefaciens* strain 17 culture, grown in M2 medium, supplemented with 0.2% cellobiose, incubated overnight at 37°C. Two microliters of the reactions

were spotted on TLC Silica gel 60F (Merck), and chromatography was carried out for 1.5 hours, using butanol:acetic acid: water 3:1:1 as a developing solvent. After drying the plate, spots were visualized by orcinol stain (5 g orcinol dissolved in 376.65 ml ethanol, 107 ml ddw and 16.15 ml sulfuric acid), and the silica plate was heated for 10 min at 70°C.

All enzymatic assays were performed at least twice in biological triplicates.

Cellulose binding assay

Binding ability of CttA to cellulose was tested by the cellulose binding assay as described earlier (81). The CBM and cohesin-CBM3a from the CipA scaffoldin of *Clostridium thermocellum* (81) were used as positive controls, and the GFP protein as a negative control for binding abilities. The binding assays was performed at least three times (biological replicates).

Comparative genomics of selected human, rumen and NHP genomes

Among the 5958 gene clusters obtained by Proteinortho, the 315 clusters common to the three groups were analyzed for verticality as well as clusters specific to one or two hosts. For verticality mapping, sequences were compared with the verticality values calculated by Nagies *et al.* (36). This was done by blasting all sequences in the database, which formed the basis for the clustering used in the latter report, against each sequence of interest. Results were filtered by an E-value of 10^{-10} , and sequences of interest were then mapped to the cluster with the highest number of hits. If the mapped cluster had a calculated verticality value, this value was then mapped to the sequence of interest.

The presence-absence of the overall 5958 gene clusters, or number of annotated glycoside hydrolases (with and without dockerin modules) obtained using dbcan2, were compared among the three groups of selected genomes (human, rumen and NHP) using PCA plot in R with phyloseq (82) and ggplot2 (83), followed by the PERMANOVA test using 1000 randomizations of the data and the vegan package (84). To highlight statistically different groups of GH, we performed a Kruskal-Wallis test, followed by false-discovery rate correction, and created abundance heatmaps for genes or transcripts, using the superheat package (<https://CRAN.R-project.org/package=superheat>) (85).

Expression of *R. hominiciens* genes in human samples

RNA was extracted in 2 positive Israeli fecal samples, using the Qiagen AllPrep PowerFecal DNA/RNA Kit, and the samples which yielded high-quality RNA were sequenced by NovaSeq SP 2x150nt (Roy J. Carver Biotechnology Center, Illinois). Reads from sample 50466110 from project PRJNA354235, which was found

positive in the MAG alignments, were also used. Reads from the metatranscriptomics of three macaque fecal samples (86) and three sheep rumen samples (59) were retrieved from the ENA database (macaque project SRX3517701-SRX3517724, samples SRR6425354, SRR6425396 and SRR6425408 and sheep project PRJNA202380, samples SRR1206249, SRR138694 and SRR138697). Reads were subsampled to 1,000,000 reads, and transcripts were quantified using RSEM (87) against their respective MAGs Human_SRR6028624_bin.16, Rumen_CACVSX01 and Macaque_bin.22. The transcripts of the annotated GHs (with and without dockerin modules) obtained with dbcan2, were compared among the three groups of selected genomes (human, rumen and NHP) using PCA plot in R with phyloseq (82) and ggplot2 (83), followed by the PERMANOVA test using 1000 randomizations of the data and the vegan package (84).

REFERENCES AND NOTES

- K. Makki, E. C. Deehan, J. Walter, F. Bäckhed, The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **23**, 705–715 (2018). doi: [10.1016/j.chom.2018.05.012](https://doi.org/10.1016/j.chom.2018.05.012); pmid: [29902436](https://pubmed.ncbi.nlm.nih.gov/29902436/)
- Y. P. Silva, A. Bernardi, R. L. Fozza, The role of short-chain fatty acids from gut Microbiota in gut-brain communication. *Front. Endocrinol. (Lausanne)* **11**, 25 (2020). doi: [10.3389/fendo.2020.00025](https://doi.org/10.3389/fendo.2020.00025); pmid: [32082260](https://pubmed.ncbi.nlm.nih.gov/32082260/)
- E. A. Bayer, H. Chanzy, R. Lamed, Y. Shoham, Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**, 548–557 (1998). doi: [10.1016/S0959-440X\(98\)80143-7](https://doi.org/10.1016/S0959-440X(98)80143-7); pmid: [9812857](https://pubmed.ncbi.nlm.nih.gov/9812857/)
- H. J. Freeman, G. A. Spiller, Y. S. Kim, A double-blind study on the effect of purified cellulose dietary fiber on 1,2-dimethylhydrazine-induced rat colonic neoplasia. *Cancer Res.* **38**, 2912–2917 (1978). pmid: [679200](https://pubmed.ncbi.nlm.nih.gov/679200/)
- S. E. Schwartz, G. D. Levine, Effects of dietary fiber on intestinal glucose absorption and glucose tolerance in rats. *Gastroenterology* **79**, 833–836 (1980). doi: [10.1016/0016-5085\(80\)90438-2](https://doi.org/10.1016/0016-5085(80)90438-2); pmid: [6252089](https://pubmed.ncbi.nlm.nih.gov/6252089/)
- S. Morais, I. Mizrachi, *Islands in the stream: From individual to communal fiber degradation in the rumen ecosystem (FEMS Reviews in Microbiology, 2019)*. I. Mizrachi, in *The Prokaryotes: Prokaryotic Biology and Symbiotic Associations*, E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, F. Thompson, Eds. (Springer, 2013), pp. 533–544.
- C. Robert, A. Bernalier-Donadille, The cellulolytic microflora of the human colon: Evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects. *FEMS Microbiol. Ecol.* **46**, 81–89 (2003). doi: [10.1016/S0168-6496\(03\)00207-1](https://doi.org/10.1016/S0168-6496(03)00207-1); pmid: [14622920.12868](https://pubmed.ncbi.nlm.nih.gov/14622920.12868/)
- C. Chassard, E. Delmas, C. Robert, P. A. Lawson, A. Bernalier-Donadille, Ruminococcus champanellensis sp. nov., a cellulose-degrading bacterium from human gut microbiota. *Int. J. Syst. Evol. Microbiol.* **62**, 138–143 (2012). doi: [10.1099/ijss.0.027375-0](https://doi.org/10.1099/ijss.0.027375-0); pmid: [21357460](https://pubmed.ncbi.nlm.nih.gov/21357460/)
- Y. Ben David *et al.*, Ruminococcal cellulosome systems from rumen to human. *Environ. Microbiol.* **17**, 3407–3426 (2015). doi: [10.1111/1462-2920.12868](https://doi.org/10.1111/1462-2920.12868); pmid: [25845888](https://pubmed.ncbi.nlm.nih.gov/25845888/)
- S. Morais *et al.*, Enzymatic function of cellulosomal enzymes from the human gut bacterium, Ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition. *Environ. Microbiol.* **18**, 542–556 (2016). doi: [10.1111/1462-2920.13047](https://doi.org/10.1111/1462-2920.13047); pmid: [26347002](https://pubmed.ncbi.nlm.nih.gov/26347002/)
- L. Artzi, E. A. Bayer, S. Morais, Cellulosomes: Bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95 (2017). doi: [10.1038/nrmicro.2016.164](https://doi.org/10.1038/nrmicro.2016.164); pmid: [27941816](https://pubmed.ncbi.nlm.nih.gov/27941816/)
- J. L. Slavin, P. M. Brauer, J. A. Marlett, Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects. *J. Nutr.* **111**, 287–297 (1981). doi: [10.1093/jn/111.2.287](https://doi.org/10.1093/jn/111.2.287); pmid: [6257867](https://pubmed.ncbi.nlm.nih.gov/6257867/)
- K. J. Wedekind, H. R. Mansfield, L. Montgomery, Enumeration and isolation of cellulolytic and hemicellulolytic bacteria from human feces. *Appl. Environ. Microbiol.* **54**, 1530–1535 (1988). doi: [10.1128/aem.54.6.1530-1535.1988](https://doi.org/10.1128/aem.54.6.1530-1535.1988); pmid: [3415224](https://pubmed.ncbi.nlm.nih.gov/3415224/)
- S. Jindou *et al.*, Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*. *FEMS Microbiol. Lett.* **285**, 188–194 (2008). doi: [10.1111/j.1574-6968.2008.01234.x](https://doi.org/10.1111/j.1574-6968.2008.01234.x); pmid: [18564339](https://pubmed.ncbi.nlm.nih.gov/18564339/)
- J. M. Brule *et al.*, Cellulosomics, a gene-centric approach to investigating the intraspecific diversity and adaptation of *Ruminococcus flavefaciens* within the rumen. *PLOS ONE* **6**, e25329 (2011). doi: [10.1371/journal.pone.0025329](https://doi.org/10.1371/journal.pone.0025329); pmid: [22043282](https://pubmed.ncbi.nlm.nih.gov/22043282/)
- R. D. Stewart *et al.*, Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019). doi: [10.1038/s41587-019-0202-3](https://doi.org/10.1038/s41587-019-0202-3); pmid: [31375809](https://pubmed.ncbi.nlm.nih.gov/31375809/)
- A. Almeida *et al.*, A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019). doi: [10.1038/s41586-019-0965-1](https://doi.org/10.1038/s41586-019-0965-1); pmid: [30745586](https://pubmed.ncbi.nlm.nih.gov/30745586/)
- D. H. Parks, M. Melfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015). doi: [10.1101/gr.186072.114](https://doi.org/10.1101/gr.186072.114); pmid: [25977477](https://pubmed.ncbi.nlm.nih.gov/25977477/)
- V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**(D1), D490–D495 (2014). doi: [10.1093/nar/gkt1178](https://doi.org/10.1093/nar/gkt1178); pmid: [24270786](https://pubmed.ncbi.nlm.nih.gov/24270786/)
- M. T. Rincon *et al.*, A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of *Ruminococcus flavefaciens*. *J. Bacteriol.* **189**, 4774–4783 (2007). doi: [10.1128/JB.00143-07](https://doi.org/10.1128/JB.00143-07); pmid: [17468247](https://pubmed.ncbi.nlm.nih.gov/17468247/)
- L. G. Wayne, D. J. Brenner, R. R. Colwell, P. A. Grimont, Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *J. Syst. Palaeontol.* **37**, 4 (1987). doi: [10.1128/JB.00143-07](https://doi.org/10.1128/JB.00143-07); pmid: [17468247](https://pubmed.ncbi.nlm.nih.gov/17468247/)
- W. B. Whitman *et al.*, Development of the SeqCode: A proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. *Syst. Appl. Microbiol.* **45**, 126305 (2022). doi: [10.1016/j.symp.2022.126305](https://doi.org/10.1016/j.symp.2022.126305); pmid: [36049255](https://pubmed.ncbi.nlm.nih.gov/36049255/)
- T. C. A. Hitch *et al.*, Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Commun.* **1**, 16 (2021). doi: [10.1038/s43705-021-00017-z](https://doi.org/10.1038/s43705-021-00017-z); pmid: [36732617](https://pubmed.ncbi.nlm.nih.gov/36732617/)
- M. C. Wibowo *et al.*, Reconstruction of ancient microbial genomes from the human gut. *Nature* **594**, 234–239 (2021). doi: [10.1038/s41586-021-03532-0](https://doi.org/10.1038/s41586-021-03532-0); pmid: [33981035](https://pubmed.ncbi.nlm.nih.gov/33981035/)
- D. K. Jacobson *et al.*, Publisher Correction: Analysis of global human gut metagenomes shows that metabolic resilience potential for short-chain fatty acid production is strongly influenced by lifestyle. *Sci. Rep.* **11**, 10114 (2021). doi: [10.1038/s41598-021-89719-x](https://doi.org/10.1038/s41598-021-89719-x); pmid: [33958685](https://pubmed.ncbi.nlm.nih.gov/33958685/)
- S. Rampelli *et al.*, Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota. *Curr. Biol.* **25**, 1682–1693 (2015). doi: [10.1016/j.cub.2015.04.055](https://doi.org/10.1016/j.cub.2015.04.055); pmid: [25981789](https://pubmed.ncbi.nlm.nih.gov/25981789/)
- M. P. Francino, Antibiotics and the human gut microbiome: Dysbioses and accumulation of resistances. *Front. Microbiol.* **6**, 1543 (2016). doi: [10.3389/fmicb.2015.01738](https://doi.org/10.3389/fmicb.2015.01738)
- J. G. Sanders *et al.*, Widespread extinctions of co-diversified primate gut bacterial symbionts from humans. *Nat. Microbiol.* **8**, 1039–1050 (2023). doi: [10.1038/s41564-023-01388-w](https://doi.org/10.1038/s41564-023-01388-w); pmid: [37169918](https://pubmed.ncbi.nlm.nih.gov/37169918/)
- K. R. Amato *et al.*, Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J.* **13**, 576–587 (2019). doi: [10.1038/s41396-018-01750-p](https://doi.org/10.1038/s41396-018-01750-p); pmid: [29995839](https://pubmed.ncbi.nlm.nih.gov/29995839/)
- M. Lechner *et al.*, Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011). doi: [10.1186/1471-2105-12-124](https://doi.org/10.1186/1471-2105-12-124); pmid: [21526987](https://pubmed.ncbi.nlm.nih.gov/21526987/)
- Y. Barak *et al.*, Matching fusion protein systems for affinity analysis of two interacting families of proteins: The cohesin-dockerin interaction. *J. Mol. Recognit.* **18**, 491–501 (2005). doi: [10.1002/jmr.749](https://doi.org/10.1002/jmr.749); pmid: [16167300](https://pubmed.ncbi.nlm.nih.gov/16167300/)
- B. Dassa *et al.*, Rumen cellulosomics: Divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PLOS ONE* **9**, e99221 (2014). doi: [10.1371/journal.pone.0099221](https://doi.org/10.1371/journal.pone.0099221); pmid: [24992679](https://pubmed.ncbi.nlm.nih.gov/24992679/)
- M. T. Rincon *et al.*, Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J. Bacteriol.* **187**, 7569–7578 (2005). doi: [10.1128/JB.187.22.7569-7578.2005](https://doi.org/10.1128/JB.187.22.7569-7578.2005); pmid: [16267281](https://pubmed.ncbi.nlm.nih.gov/16267281/)
- O. Salama-Alber *et al.*, Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural

buttresses. *J. Biol. Chem.* **288**, 16827–16838 (2013). doi: [10.1074/jbc.M113.466672](https://doi.org/10.1074/jbc.M113.466672); pmid: [23580648](https://pubmed.ncbi.nlm.nih.gov/23580648/)

36. F. S. P. Nagies, J. Brueckner, F. D. K. Tria, W. F. Martin, A spectrum of verticality across genes. *PLOS Genet.* **16**, e1009200 (2020). doi: [10.1371/journal.pgen.1009200](https://doi.org/10.1371/journal.pgen.1009200); pmid: [33137105](https://pubmed.ncbi.nlm.nih.gov/33137105/)

37. A. Darvill, M. McNeil, P. Albersheim, D. P. Delmer, The primary cell walls of flowering plants. *The biochemistry of plants.* **1**, 91–162 (1980). doi: [10.1371/journal.pgen.1009200](https://doi.org/10.1371/journal.pgen.1009200); pmid: [33137105](https://pubmed.ncbi.nlm.nih.gov/33137105/)

88. M. Itzhak et al., Data from: Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans. Version v1, Zenodo (2024); <https://zenodo.org/records/10650136>.

ACKNOWLEDGMENTS

The authors thank D. Perlman for scientific illustrations and assistance with graphic design, A. Segal for providing human faecal samples and A. Moeller for sharing his scripts and methodologies for cospeciation analysis. **Funding:** This work was funded by grants

from the German-Israeli Project Cooperation (DIP 2476/2 -1) to I.M. and W.F.M., the European Research Council (ERC 866530 and ERC-POC 01082166) to I.M. and W.F.M. (I01018894), the Israel Science Foundation (ISF 1947/19) to I.M. and S.M. Additional funds were provided by the National Institute of Biotechnology in the Negev. **Author contributions:** S.M. supervised the research, performed biochemistry and bioinformatic analyses, analyzed the data, and cowrote the paper; S.W. helped with the metagenome samples for strain prevalence analysis; A.Z. helped with the strain prevalence and abundance analysis and provided critical insights; L.L. provided bioinformatic support and performed evolutionary analyses; F.S.P.N. analyzed verticality values of the core proteins; N.K. performed phylogenetic analyses of the core proteins; E.S.L. and A.A.F. performed cloning and biochemistry experiments; D.N.B. provided knowledge for GH98 characterization; M.P.Y. prepared corn glucuronoarabinoxylan; E.A.B. and W.F.M. provided valuable knowledge and resources, analyzed the data, and critically read the paper; and I.M. supervised the research, secured funding, designed the experiments, analyzed the data, and cowrote the paper.

Conflicts of interest: The authors declare no conflicts of interest.

Data and materials availability: Metatranscriptomes have been deposited in GenBank (PRJNA951949). Data source files are available at DRYAD (88). Corn glucuronoarabinoxylan is available from M. P. Yadav under a material transfer agreement with the USDA. **License information:** Copyright © 2024 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adj9223

Figs. S1 to S15

Tables S1 to S11

References (38–87)

Submitted 25 July 2023; accepted 8 February 2024
10.1126/science.adj9223